**PROYECTO FINAL DE MINERÍA DE DATOS**

Desarrollar un proyecto de Minería de Datos en python, siguiendo la metodología CRISP-DM con **datos** **reales** en equipos de 2 personas. Los datos deben tener mínimo 10 variables y 300 registros.

1. (1.5) Aplicar todas las fases de la metodología CRISP-DM, como ayuda se sugiere el formato adjunto. Se debe entregar un informe con todas las fases documentadas.
2. (1.0) Desarrollar un jupyter notebook de preparación de datos en Python, incluyendo todos los pasos vistos en clase. Documentar los resultados en el informe de la metodología. Incluir el pandas profiling de los datos.
3. (1.5) Crear un modelo predictivo avanzado en Python, donde:
   1. Se balancea sólo el 70% de los datos (en caso de ser necesario el balanceo)
   2. Se realiza una validación cruzada con el 70%
   3. Se aplican 4 métodos de aprendizaje supervisado de máquinas
   4. Se aplican 3 métodos de ensamble
   5. Se calculan al menos 4 medidas de calidad de cada modelo y se comparan para seleccionar los mejores modelos. Se deben interpretar todas las medidas obtenidas.
   6. De los 7 modelos creados, se seleccionan los 3 mejores. Para seleccionar los mejores modelos se debe aplicar un proceso de análisis de diferencia estadística significativa (ANOVA y Tukey).
   7. Los 3 modelos seleccionados deben pasar por un proceso de hiperparametrización con gridsearch y optimización (algoritmos genéticos/optimización bayesiana). El mejor modelo resultante se almacena para ser llevado a despliegue.
   8. El modelo final se debe almacenar en un Pipe con las operaciones de preparación de los datos para el despliegue.
   9. Se realiza un despliegue con interfaz gráfica streamlit
4. (1.0) Sustentar el desarrollo completo del trabajo en 20 minutos por equipo, mostrando los resultados de cada fase de la metodología CRISP-DM (deben participar todos los integrantes).

**Entregables**

* Documentación de la metodología CRISP-DM
* Jupyter notebook de preparación de datos. Incluir el pandas profiling.
* Jupyter notebook de creación y evaluación de modelos predictivos
* Jupyter notebook de despliegue con interfaz gráfica
* Presentar el trabajo en github